working paper
department
of economics

AN INEQUALITY FOR VECTOR-VALUED MARTINGALES
AND ITS APPLICATIONS

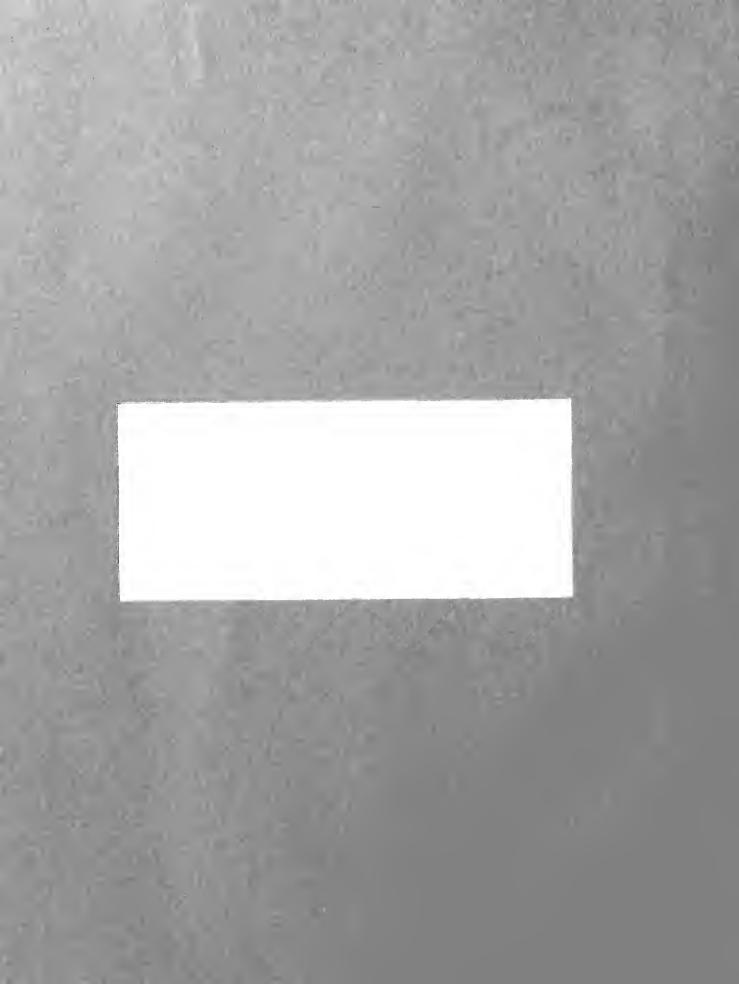Jushan Bai

96-16   June 1996

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

AN INEQUALITY FOR VECTOR-VALUED MARTINGALES
AND ITS APPLICATIONS

Jushan Bai

96-16   June 1996

An Inequality for Vector-Valued Martingales and its Applications*

Jushan Bai

Massachusetts Institute of Technology

June, 1996

Abstract

In this paper, we derive a general Hájek-Rényi type inequality for vector-valued martingales. Several well known inequalities are shown to be special cases of this general inequality. We also derive a similar inequality for dependent sequences. We then apply the inequality to the problem of strong consistency of least squares estimators for multiple regressions.

1   Introduction

For a sequence of independent random variables $\{z_i\}_{i=1}^{\infty}$ with zero mean and finite variance, Hájek and Rényi (1955) proved the following inequality: for any $\epsilon > 0$ and for any positive integers n and N $(n < N)$,

$$P\left(\max_{n \leq k \leq N} c_k |z_1 + \cdots + z_k| \leq \epsilon\right) \leq \frac{1}{\epsilon^2}\left(c_n^2 \sum_{i=1}^{n} \sigma_i^2 + \sum_{i=n+1}^{N} c_i^2 \sigma_i^2\right)$$

where c_k $(k = 1, 2, ...)$ is a sequence of non-increasing and positive numbers and $E z_i^2 = \sigma_i^2$. The Hájek-Rényi inequality is very useful in establishing strong laws of large numbers. More interestingly, this type of inequality plays an important role in the change-point problem for both parametric and nonparametric estimations [see, e.g., Bhattacharya (1987), Dumbgen (1991), and Bai (1994,1995)].

This inequality has been extended in the literature into various settings; see, e.g., Chow (1960) for martingales, Birnbaum and Marshall (1961) for continuous time martingales, and Sen (1971) for vector-valued random sequences. More recently, Christofides and Serfling (1990) extend the inequality to random fields. In this note we derive a general Hájek-Rényi type inequality for vector-valued martingales. Several special cases of this general inequality are then discussed. One of them concerns a sequence of quadratic forms of martingales constructed using a decreasing sequence of positive-semidefinite matrices (Corollary 1), as opposed to a decreasing sequence of scalars.

This quadratic form inequality is used to prove the strong consistency of least squares estimators for multiple regressions. This latter problem has received considerable attention in the literature, e.g., Anderson and Taylor (1976), Chen, Lai and Wei (1981), Drygas (1976), Lai, Robbins, and Wei (1978,1979) and Solo (1981). Using our inequality, we give an alternative and simpler proof of strong consistency.

Furthermore, we derive a similar quadratic form inequality for dependent sequences. In particular, we allow the dependence sequence to be a mixingale, which includes linear processes and strong mixing sequences as special cases. This inequality is useful in deriving the rate of convergence of change-point estimators for multiple regression models

with dependent disturbances.

2 The Inequality

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_i$ be an increasing sequence of σ-fields such that $\mathcal{F}_i \subset \mathcal{F}$. Let $Z_1, Z_2, \ldots$ be a sequence of p-dimensional martingale differences relative to $\{\mathcal{F}_i\}$ such that

$$E(Z_n | \mathcal{F}_{n-1}) = 0 \quad a.s \quad \text{and} \quad E(Z_i Z_i') = \Sigma_i.$$

Let $S_k = \sum_{i=1}^k Z_i$. Let $h_k(x)$ be a sequence of real-valued convex functions defined on R^p such that $h_1(x) \geq h_2(x) \geq \ldots \geq 0$ for all $x \in R^p$. Thus h_k is nonnegative and nonincreasing with respect to k. The general inequality is given in the following theorem:

Theorem 1 *Let $h_k(x)$ be a sequence of convex, nonnegative, and nonincreasing (with respect to k) functions defined on R^p with $E h_1(S_i) < \infty$ $(i \geq 1)$, and let Z_k be a sequence of R^p-valued martingale differences relative to an increasing sequence of σ-fields $\{\mathcal{F}_k\}$. Then for any $\epsilon > 0$ and for any positive integers n and N $(n < N)$, we have*

$$(1) \qquad P\left(\max_{n \leq k \leq N} h_k(S_k) \geq \epsilon\right) \leq \frac{1}{\epsilon}\left(E[h_n(S_n)] + \sum_{i=n+1}^{N} E[h_i(S_i) - h_i(S_{i-1})]\right).$$

When $p = 1$ (the one dimensional case), if we let $h(x) = x^2$, we obtain the Kolmogorov type inequality. Let c_k be a sequence of nonincreasing and positive numbers. If we choose $h(x) = c_k^2 x^2$, then (1) specializes to the inequality of Hájek and Rényi (1955) (for independent random variables). Further, let $h_k(x) = c_k x^+ = c_k \max\{0, x\}$, then Chow's inequality is obtained. For $p \geq 1$, let $h_k(x) = c_k \sup_{\lambda \neq 0} (\lambda' A \lambda)^{-1/2} |\lambda' x|$, where A is positive-definite. Then $h_k(x)$ is convex and nonincreasing with respect to k. For this sequence of h_k, (1) yields the inequality of Sen (1971).

We now consider another special case of the theorem, which is useful in establishing the strong consistency of least squares estimators. Let M_k $(k = 1, 2, \ldots)$ be a sequence of positive-semidefinite matrices $(p \times p)$ and nonincreasing in the sense that $M_k - M_{k+1}$ is positive-semidefinite. Let

$$(2) \qquad h_k(x) = x' M_k x.$$

3

Clearly, $h_k(x)$ is convex, nonnegative, and nonincreasing with respect to k. We have

Corollary 1 *Let M_k be a sequence of $p \times p$ positive-semidefinite and nonincreasing matrices, and let Z_k be given in Theorem 1, then*

$$P\left(\max_{n \leq k \leq N} S_k' M_k S_k \geq \epsilon^2\right) \leq \frac{1}{\epsilon^2}\left(tr\left(M_n \sum_{i=1}^{n} \Sigma_i\right) + \sum_{i=n+1}^{N} tr(M_i \Sigma_i)\right)$$

where $tr(\cdot)$ is the trace function.

Proof of Corollary 1: We only need to compute the right hand side of (1) for this particular h_k. Now $E[h_n(S_n)] = E(S_n' M_n S_n) = tr(M_n E[S_n S_n']) = tr(M_n \sum_{i=1}^{n} \Sigma_i)$ and

$$E[h_k(S_k) - h_k(S_{k-1})] = E[S_k' M_k S_k - S_{k-1}' M_k S_{k-1}]$$

$$= E(Z_k' M_k Z_k) - 2E[S_{k-1}' M_k Z_k]$$

$$= tr(M_k E[Z_k Z_k']) = tr(M_k \Sigma_k).$$

We have used the fact that $E[S_{k-1}' M_k Z_k] = E[S_{k-1} M_k E(Z_k|\mathcal{F}_{k-1})] = 0$. This proves the corollary. $\square$

Remark 1. Because $h_k(x) = c_k \sup_{\lambda \neq 0}(\lambda' A \lambda)^{-1/2}|\lambda' x| = c_k(x' A^{-1} x)^{1/2}$, where A is positive-definite, we see that Corollary 1 extends the inequality of Sen (1971). The nonincreasingness of Sen's h_k (with respect to k) is due to the nonincreasingness of the scalar sequence c_k, whereas the nonincreasingness of h_k defined in (2) is due to the nonincreasingness of the matrix sequence M_k. It is this latter feature that makes the result of Corollary 1 suitable for the least squares problem mentioned in the Introduction. $\square$

Proof of Theorem 1: The proof is similar to that of Chow (1960). Let A be the event $\{\sup_{n \leq k \leq N} h_k(S_k) \geq \epsilon\}$. Define $A_k = \{h_r(S_r) < \epsilon \text{ for } n \leq r < k; h_k(S_k) \geq \epsilon\}$. Then the A_k are disjoint, $A_k \in \mathcal{F}_k$ and $\cup_{i=n}^{N} A_k = A$. Thus

$$\epsilon P(A) = \epsilon \sum_{k=n}^{N} P(A_k) \leq \sum_{k=n}^{N} \int_{A_k} h_k(S_k) dP$$

$$\leq E h_n(S_n) - \int_{\Omega - A_n} h_n(S_n) dP + \int_{A_{n+1}} h_{n+1}(S_{n+1}) dP + \sum_{n+2}^{N} \int_{A_k} h_k(S_k) dP$$

4

$$\leq Eh_n(S_n) + \int_{\Omega - A_n} [h_{n+1}(S_{n+1}) - h_{n+1}(S_n)]dP$$

$$- \int_{\Omega - (A_n \cup A_{n+1})} h_{n+1}(S_{n+1})dP + \sum_{n+2}^{N} \int_{A_k} h_k(S_k)dP$$

The last inequality follows from $h_n(S_n) \geq h_{n+1}(S_n)$. A key observation is that for each fixed k, the sequence $\xi_i = h_k(S_i)$ $(i = 1, 2, ...)$ is a real-valued submartingale relative to $\{\mathcal{F}_i\}$ because $h_k(\cdot)$ is convex and S_i is a martingale [Doob (1953), p.295]. Thus

(3)
$$\int_{A_n} [h_{n+1}(S_{n+1}) - h_{n+1}(S_n)]dP \geq 0$$

and more generally

(4)
$$\int_{A_n \cup \cdots \cup A_k} [h_{k+1}(S_{k+1}) - h_{k+1}(S_k)]dP \geq 0 \quad (k > n).$$

It follows that

$$\epsilon P(A) \leq Eh_n(S_n) + E[h_{n+1}(S_{n+1}) - h_{n+1}(S_n)]$$

$$+ \int_{\Omega - (A_n \cup A_{n+1})} [h_{n+2}(S_{n+2}) - h_{n+2}(S_{n+1})]dP$$

$$- \int_{\Omega - (A_n \cup A_{n+1} \cup A_{n+2})} h_{n+2}(S_{n+2}) + \sum_{n+3}^{N} \int_{A_k} h_k(S_k)dP.$$

The theorem follows by repeating the above argument and making use of (4). $\square$.

Next, we generalize the inequality to dependent sequences. Let $\{\mathcal{F}_i\}_{i=-\infty}^{\infty}$ be a sequence of increasing σ-fields. Assume that $\{Z_i, \mathcal{F}_i\}$ forms an L^2-mixingale sequence (Hall and Heyde 1980, p. 21). That is, there exist nonnegative constants $\{b_i : i \geq 1\}$ and $\{\psi_m : m \geq 0\}$ such that $\psi_m \downarrow 0$ as $m \to \infty$ and for all $i \geq 0$ and $m \geq 0$, and

(i) $E\|E(Z_i|\mathcal{F}_{i-m})\|^2 \leq b_i^2 \psi_m^2$,

(ii) $E\|Z_i - E(Z_i|\mathcal{F}_{i+m})\|^2 \leq b_i^2 \psi_{m+1}^2$,

where $\|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ for $x \in R^p$. We assume in addition that

(iii) $\sum_{m=1}^{\infty} m^{1+\delta} \psi_m < \infty$, for some $\delta > 0$.

Throughout the remainder of this paper, any L^2 mixingale $\{Z_i, \mathcal{F}_i\}$ is implicitly assumed to satisfy conditions (i)-(iii). The sequence b_i is called mixing norms and ϕ_i

is called mixing coefficient. Mixingales include martingale differences, linear processes, and strong mixing processes as special cases. Many other dependent structures are also special cases of mixingales; see Andrews (1988). For example, let $Z_n = \sum_{j=-\infty}^{\infty} d_j \eta_{n-j}$ with $\sum_{j=-\infty}^{\infty} d_j^2 < \infty$ and η_j i.i.d. mean zero and finite variance σ^2. Let $\mathcal{F}_n = \sigma$-field$\{\eta_j; j \leq n\}$. Then $\{Z_i, \mathcal{F}_i\}$ is a mixingale with $b_i^2 = \sigma^2$ (for all i) and $\psi_i^2 = \sum_{|k| \geq i} d_k^2$.

Theorem 2 *Let M_k be a sequence of $p \times p$ non-random positive-semidefinite and non-increasing matrices, and let Z_k be an L^2 mixingale. Then, for every $\epsilon > 0$,*

$$(5) \qquad P\left(\max_{n \leq k \leq N} S_k' M_k S_k \geq \epsilon^2\right) \leq \frac{C}{\epsilon^2}\left(tr(M_n)\sum_{\ell=1}^{n} b_\ell^2 + \sum_{i=n+1}^{N} b_i^2\, tr(M_i)\right).$$

where $C = 4(\psi_0^2 + 2\sum_{j=1}^{\infty} j^{2+2\delta}\psi_j^2)(1 + 2\sum_{j=1}^{\infty} j^{-2-2\delta}) < \infty$.

Proof of Theorem 2: We can write (Hall and Heyde 1981, p20),

$$Z_k = \sum_{j=-\infty}^{\infty} Z_{jk}, \quad \text{with } Z_{jk} = E(Z_k|\mathcal{F}_{k-j}) - E(Z_k|\mathcal{F}_{k-j-1}).$$

Thus $M_k^{1/2}S_k = M_k^{1/2}\sum_{i=1}^{k} Z_i = \sum_{j=-\infty}^{\infty} M_k^{1/2}\sum_{i=1}^{k} Z_{ji} = \sum_{j=-\infty}^{\infty} M_k^{1/2}S_{jk}$, where $S_{jk} = \sum_{i=1}^{k} Z_{ji}$. Thus

$$P\left(\max_{n \leq k \leq N}\|M_k^{1/2}S_k\| \geq \epsilon\right) \leq P\left(\sum_{j=-\infty}^{\infty}\max_{n \leq k \leq N}\|M_k^{1/2}S_{jk}\| \geq \epsilon\right).$$

Note that for each j, $\{S_{jk}, \mathcal{F}_{k-j}; 1 \leq k \leq N\}$ is a martingale. Let $a_j > 0$ for all j such that $\sum_{j=-\infty}^{\infty} a_j = 1$. Then

$$(6) \qquad P\left(\sum_{j=-\infty}^{\infty}\max_{n \leq k \leq N}\|M_k^{1/2}S_{jk}\| \geq \epsilon\right) \leq \sum_{j=-\infty}^{\infty} P\left(\max_{n \leq k \leq N}\|M_k^{1/2}S_{jk}\| \geq a_j\epsilon\right)$$

$$\leq \frac{1}{\epsilon^2}\sum_{j=-\infty}^{\infty} a_j^{-2}\left(tr\left(M_n\sum_{i=1}^{n}\Sigma_{ji}\right) + \sum_{i=n+1}^{N} tr(M_i\Sigma_{ji})\right)$$

where $\Sigma_{ji} = E(Z_{ji}Z_{ji}')$ and the second inequality follows from Corollary 1. For any vector $x \in R^p$, we have $x'xI - xx' \geq 0$ (positive-semidefinite), where I is a $p \times p$ identity matrix. Substitute Z_{ji} for x and take expectation to obtain $E\|Z_{ji}\|^2 I - \Sigma_{ji} \geq 0$. Furthermore, by

the mixingale property and the definition of Z_{ji}, it is easy to show that $E\|Z_{ji}\|^2 \le 4b_i^2\psi_{|j|}^2$ for all j. Thus $4b_i^2\psi_{|j|}^2 I - \Sigma_{ji} \ge 0$. Using the fact that $tr(CA) \le tr(CB)$ for $B \ge A$ and $C \ge 0$, we have $tr(M_\ell\Sigma_{ji}) \le 4b_i^2\psi_{|j|}^2 tr(M_\ell)$. It follows that the right hand side of (5) is bounded by

$$\frac{4}{\epsilon^2}\left(\sum_{j=-\infty}^\infty a_j^{-2}\psi_{|j|}^2\right)\left(tr\Big(M_n\Big)\sum_{i=1}^n b_i^2 + \sum_{i=n+1}^N b_i^2\, tr\Big(M_i\Big)\right).$$

It remains to choose appropriate $a_j's$ such that $\sum_j a_j^{-2}\psi_{|j|}^2$ is bounded. Let $\nu_0 = 1$ and $\nu_j = j^{-1-\delta}$ $(j \ge 1)$, where $\delta > 0$ as given in (iii). Let $a_j = \nu_j/(1+2\sum_{k=1}^\infty \nu_k)$ and $a_{-j} = a_j$ for $j \ge 0$. Then $\sum_j a_j = 1$. By assumption (iii), $\sum_j a_j^{-2}\psi_{|j|}^2 = (\psi_0^2 + 2\sum_{j=1}^\infty j^{2+2\delta}\psi_j^2)(1 + 2\sum_{j=1}^\infty j^{-2-2\delta}) < \infty$. $\qquad \square$

Remark 2. Inspecting the proof we see that if $\Sigma_{ji} \le \Sigma_i\psi_{|j|}^2$, for some positive-semidefinite Σ_i, then from $tr(M_\ell\Sigma_{ji}) \le tr(M_\ell\Sigma_i)\psi_{|j|}^2$ and (6), we have

$$(7) \qquad P\left(\max_{n\le k\le N} S_k'M_kS_k \ge \epsilon^2\right) \le \frac{C}{\epsilon^2}\left(tr\Big(M_n\sum_{i=1}^n \Sigma_i\Big) + \sum_{i=n+1}^N tr(M_i\Sigma_i)\right),$$

which is analogous to Corollary 1. This situation arises in the following setup: $Z_i = x_i\varepsilon_i$, where x_i is a $p \times 1$ vector of nonrandom constants and ε_i is random variable such that $\{\varepsilon_i, \mathcal{F}_i\}$ forms an L^2 mixingale with mixing norms $\{b_i^2\}$ and mixing coefficients $\{\psi_m^2\}$. We can write $\varepsilon_i = \sum_{j=-\infty}^\infty \varepsilon_{ji}$, so that $Z_i = \sum_{j=-\infty}^\infty Z_{ji}$ with $Z_{ji} = x_i\varepsilon_{ji}$. It follows that $\Sigma_{ji} = E(Z_{ji}Z_{ji}') = x_ix_i'E(\varepsilon_{ji}^2) \le (x_ix_i')4b_i^2\psi_{|j|}^2$. We can choose $\Sigma_i = (4b_i^2)x_ix_i'$. This setup is related to regression models, where x_i are regressors and ε_i are disturbances. Inequality (7) will be used in proving strong consistency of least squares estimators in regression models with mixingale disturbances. $\qquad \square$

Corollary 2 *For Z_i in Theorem 2, assume that the mixing norms b_i^2 are such that $\sup_i b_i^2 \le b < \infty$. Then, for some $M < \infty$,*

$$P\left(\max_{n\le k\le N}\frac{1}{k}\|\sum_{i=1}^k Z_i\| \ge \epsilon\right) \le \frac{M}{\epsilon^2 n} \qquad \text{for all } N \ge n \ge 1.$$

Proof: Let $M_k = k^{-2}I$. Then $tr(M_n)\sum_{i=1}^n b_i^2 \le pb^2/n$ and $\sum_{i=n+1}^N b_i^2 tr(M_i) \le p\sum_{i=n+1}^\infty b^2/i^2 = O(1/n)$. Thus the right hand side of (5) is bounded by $M/(\epsilon^2 n)$ for some $M < \infty$. $\qquad \square$

The corollary is useful in change-point problems. Note that since the probability bound does not depend on N, the corollary holds when we replace N by ∞. This yields the strong law of large numbers for mixingales as well as certain rate of convergence for the strong law.

3   Application

The application concerns the strong consistency of least squares estimator of the parameter β_j in the following multiple regression:

$$(8) \qquad y_i = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + \varepsilon_i = x_i'\beta + \varepsilon_i \quad (i = 1, 2, \ldots)$$

where the vector $x_i = (x_{i1}, \ldots, x_{ip})'$ $(i \geq 1)$ consists of known constants, $\beta = (\beta_1, \ldots, \beta_p)'$ is the regression coefficient, and $\{\varepsilon_i\}$ is an L^2 mixingale with respect to an increasing sequence of σ-fields $\{\mathcal{F}_i\}$. Denote by $X_n = (x_1, \ldots, x_n)'$ the design matrix and let $Y_n = (y_1, \ldots, y_n)'$ and $U_n = (\varepsilon_1, \ldots, \varepsilon_n)'$. Assume $(X_n'X_n)$ is positive-definite for some $n = m$ and hence for all $n > m$. The least squares estimator b_n, based on the first n observations, is given by

$$b_n = (X_n'X_n)^{-1}X_n'Y_n = \beta + (X_n'X_n)^{-1}X_n'U_n.$$

Thus strong consistency is equivalent to

$$(9) \qquad (X_n'X_n)^{-1}X_n'U_n \to 0 \quad \text{with probability 1.}$$

When the regressors are non-random, Lai, Robbins, and Wei (1977, 1979) establish the strong consistency under the minimal condition

$$(10) \qquad (X_n'X_n)^{-1} \to 0$$

together with the assumption that ε_i is a martingale-difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_i\}$ such that $\sup_i E(\varepsilon_i^2|\mathcal{F}_{i-1}) < \infty$ (a.s.). Condition (10) is significantly weaker than the conventional assumption that $(X_n'X_n/n)$ converges to a positive definite matrix. Under the latter assumption, the condition number of

$(X'_n X_n)$ (the ratio of the largest and smallest eigenvalues) will be uniformly bounded, and the strong consistency will be trivial to prove. Condition (9) is known as a necessary and sufficient condition when the ε_i are i.i.d. with zero mean and finite variance; see Lai et al. (1979). For dynamic models, strong consistency is also studied by many authors (cf. Anderson and Taylor, 1979, Chri Christopeit and Helmes, 1980, Lai and Wei, 1982).

Using our inequality, we give an alternative and simpler proof of strong consistency. We do impose a stronger condition than (9), but we allow the disturbances to have broad correlation structures. Let $\lambda_{\max}(n)$ denote the largest eigenvalue of $X'_n X_n$ and $\lambda_{\min}(n)$ the minimum eigenvalue of $X'_n X_n$. Let Φ be the set of functions $\phi(x)$ such that each $\phi(x)$ is positive and non-decreasing in the interval $x > x_0$ for some x_0 and the series $\sum \frac{1}{n\phi(n^\alpha)}$ converges for every $\alpha > 0$. For example $\phi(x) = x^\delta$, $\phi(x) = [\log(x)]^{1+\delta}$ and $\phi = \log(x)[\log\log(x)]^{1+\delta}$ ($\delta > 0$) are members of Φ. We assume, for some $\phi \in \Phi$,

$$(11) \qquad \lambda_{\min}(n) \to \infty \quad and \quad \phi(\lambda_{\max}(n)) = O(\lambda_{\min}(n)).$$

This condition is weaker than (2.24) of Lai and Wei (1982) but slightly stronger than their (1.17). For dynamic models, Lai and Wei show that condition (11) is close to be weakest possible for strong consistency.

Next, we assume $\{\varepsilon_i\}$ is an L^2 mixingale with respect to an increasing sequence of σ-fields $\{\mathcal{F}_i\}$. Further assume that $\sup_i b_i^2 < b^2 < \infty$. When ε_i is a martingale-difference sequence, b_i^2 can be taken as $E\varepsilon_i^2$. Note that $\sup_i E\varepsilon_i^2 < \infty$ can be significantly weaker than $\sup_i E(\varepsilon_i^2|\mathcal{F}_{i-1}) < \infty$ (a.s.), as is assumed in the previous literature. For example, let $\varepsilon_i = \eta_i\eta_{i-1}$, where η_i are i.i.d. normal. Let $\mathcal{F}_i = \sigma$-field$\{\eta_j; j \le i\}$. Then ε_i is a sequence of martingale differences. The condition that $\sup_i E(\varepsilon_i^2|\mathcal{F}_{i-1}) < \infty$ is equivalent to $\sup_i \eta_i^2 < \infty$, which does not hold.

Theorem 3 *Let $\{\varepsilon_i, \mathcal{F}_i\}$ be an L^2 mixingale with $\sup_i b_i^2 \le b^2 < \infty$. Assume (11) holds. Then the least squares estimator b_n is strongly consistent for β.*

Proof. We observe that

$$\|b_k - \beta\|^2 = \|(X'_k X_k)^{-1} X'_k U_k\|^2$$

9

$$\text{(12)} \qquad \le \quad \|(X_k'X_k)^{-1/2}\|^2 \|(X_k'X_k)^{-1/2}X_kU_k\|^2$$

$$= \quad (\lambda_{\min}(k))^{-1}\{U_k'X_k(X_k'X_k)^{-1}X_k'U_k\}$$

Let $Z_i = x_i\varepsilon_i$. Then Z_i is a R^p-valued mixingale with respect to $\{\mathcal{F}_i\}$. Let $S_k = \sum_{i=1}^{k} x_i\varepsilon_i = X_k'U_k$, and let $M_k = c_k^{-1}(X_k'X_k)^{-1}$ with $c_k = \lambda_{\min}(k)$. We see that M_k is non-increasing, and $\|b_k - \beta\|^2 \le S_k'M_kS_k$ by (12). Thus by Remark 2,

$$\text{(13)} \quad P\Big(\sup_{n \le k \le N} \|b_k - \beta\|^2 > \epsilon^2\Big) \quad \le \quad P\Big(\sup_{n \le k \le N} S_k'M_kS_k > \epsilon^2\Big)$$

$$\le \quad \frac{C}{\epsilon^2}\Big(b^2\, tr\Big(M_n \sum_{i=1}^{n} \Sigma_i\Big) + b^2 \sum_{i=n+1}^{N} tr(M_i\Sigma_i)\Big)$$

where $\Sigma_i = x_ix_i'$. Note that

$$\text{(14)} \qquad tr\Big(M_n \sum_{i=1}^{n} \Sigma_i\Big) = c_n^{-1}tr\Big((X_n'X_n)^{-1}X_n'X_n\Big) = pc_n^{-1} = p/\lambda_{\min}(n).$$

Furthermore,

$$tr(M_i\Sigma_i) = x_i'M_ix_i = c_i^{-1}x_i(X_i'X_i)^{-1}x_i = c_i^{-1}(|A_i| - |A_{i-1}|)/|A_i|$$

where $A_i = (X_i'X_i)$ and $|A_i|$ is the determinant of A_i. Because $|A_n| \le \lambda_{\max}(n)^p$, or equivalently, $|A_n|^{1/p} \le \lambda_{\max}(n)$, we have $\phi(|A_n|^{1/p}) \le \phi(\lambda_{\max}(n)) = O(\lambda_{\min}(n))$ by (11). This implies that

$$c_i^{-1}(|A_i| - |A_{i-1}|)/|A_i| \le L\frac{(|A_i| - |A_{i-1}|)}{|A_i|\phi(|A_i|^{1/p})}$$

for some $L < \infty$. Thus

$$\text{(15)} \qquad \sum_{i=n+1}^{N} tr(M_i\Sigma_i) = \sum_{i=n+1}^{N} c_i^{-1}(|A_i| - |A_{i-1}|)/|A_i| \le L \sum_{i=n+1}^{\infty} \frac{(|A_i| - |A_{i-1}|)}{|A_i|\phi(|A_i|^{1/p})}.$$

Because $\sum 1/(n\phi(n^{1/p})) < \infty$, the above series converges by the integral comparison test. Combining (13)- (15), and letting $N \to \infty$, we have

$$P\left(\sup_{k \ge n} \|b_k - \beta\|^2 > \epsilon^2\right) \le \frac{C}{\epsilon^2}\left(\frac{pb^2}{\lambda_{\min}(n)} + b^2L \sum_{i=n+1}^{\infty} \frac{(|A_i| - |A_{i-1}|)}{|A_i|\phi(|A_i|^{1/p})}\right).$$

The right hand side above converges to zero as $n \to \infty$. $\quad\square$

References

[1] Anderson, T.W. and Taylor J.B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* 4 788-790.

[2] Anderson, T.W. and Taylor J.B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* 7 484-489.

[3] Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* 15 453-472.

[4] Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory* 11 403-436.

[5] Bhattacharya, P.K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General Multiparameter case. *Journal of Multivariate Analysis* 23 183-208.

[6] Birnbaum, Z.W. and Marshall A.W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Stat.* 32 687-703.

[7] Chen, G.J., Lai, T.L., and Wei, C.Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* 11 319-333.

[8] Chow, Y.S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc,* 11 107-111.

[9] Christopeit, N. and Helmes, N.C. (1980). Strong consistency of least squares estimators in linear regression models. *Ann. Statist.* 8 778-788.

[10] Christofides, T.C. and Serfling, R.J. (1990). Maximal inequalities for multidimensionally indexed submartingale arrays. *Ann. Probab.* 18 630-641.

11

[11] Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrschein. verw. Gebiete.* 34 119-127.

[12] Dumbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* 19 1471-1495.

[13] Hájek, J. and Rényi A. (1955). Generalization of an inequality of Kolmogorov. *Acta. Math. Acad. Sci. Hungar.* 6 281-283.

[14] Lai, T.L., Robbins H. and Wei C.Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. USA* 75 3034-3036.

[15] Lai, T.L., Robbins H. and Wei C.Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multi. Analysis* 9 343-361.

[16] Lai, T.L. and Wei C.Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* 10 154-166.

[17] Sen, P.K. (1971). A Hájek-Rényi type inequality for stochastic vectors with applications to simultaneous confidence regions. *Ann. Math. Statist.* 42 1132-1134.

[18] Solo, V. (1981). Strong consistency of least squares estimators in regression with correlated disturbances. *Ann. Statist.* 9 689-693.

Date Due